

B.U.-68-M
R. P. Steel

Systems of elements. Suppose we have an aggregate, a class, a collection or a system of undefined objects. Label these with the letters of the alphabet, with a letter and subscripts, or in some other convenient manner. E.g., a, b, ... or X_1, X_2, \dots .

We shall have the real numbers in mind most often and will be concerned with practical problems in connection with their combination. A system could consist of the positive integers, the rational numbers.

Operations. A rule or rules of combination may be defined for our system.

For the real numbers, we shall be concerned with the understood operations of addition (+) and multiplication (x) and their inverses, subtraction (-) and division (\div). Division by zero is not permitted.

These rules, for the system of real numbers, possess certain properties which do not hold for all systems and operations. The properties are:

	Addition	Multiplication
Commutativity	$a+b = b+a$	$ab = ba$
Associativity	$a+(b+c) = (a+b)+c$	$a(bc) = (ab)c$

The game. We now proceed to use our system and the rules to play a game of solitaire. The end result will be given and the player required to use the system and rules to attain this result. No inferences will be required. The process is deduction, not induction.

Notation. Let X denote a variable. If n observations are made on X, denote them by X_1, X_2, \dots, X_n . For example, X might denote weight of children at birth. For $n = 4$ children, we might observe $X_1 = 7$ lbs. 3 oz., $X_2 = 8$ lbs. 1 oz., $X_3 = 7$ lbs. 10 oz. and $X_4 = 6$ lbs. 15 oz. The subscripts serve as labels or names for the individual weights. These have been given in the order in which the observations were made and not in a rank order.

A common operation in statistics is addition. We may write

$$X_1 + X_2 + \dots + X_i + \dots + X_n,$$

which will certainly be a nuisance, or develop a shorthand. A common shorthand is the left-hand side of the following equation:

$$\sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_i + \dots + X_n.$$

The right-hand side is essentially a definition of the left. Σ denotes summation. We sum all the X 's or, alternately, the X_i 's for $i = 1, 2, \dots, n$. The first subscript is written below the Σ and the last above. We term i an index, an index of summation, and it takes on integral values. In the longhand form, the first and second subscripts are generally shown to set the pattern of subscripts. E.g., $X_1 + X_3 + \dots$ indicates that every other observation, beginning with the first is to be added.

PROBLEMS OF NOTATION

Given a set of observations $X_1, X_2, \dots, X_i, \dots, X_n$ which are the numbers 17, 21, 13, 19, 19, 20, 18, 16.

1. What is the value of n ?
2. " " " " " X_i for $i = 1? 3? 6?$
3. " " " " " i " $X_i = 17? 18? 19?$
4. " " " " " X_{2i} " $i = 1? 2? 3?$
5. " " " " " X_{2i+1} for $i = 1? 3?$

PROBLEMS involving addition

Using the numbers given for the previous set of questions:

1. Find $\sum_{i=1}^n X_i$, $\sum_{i=1}^4 X_{2i}$, $\sum_{i=1}^4 X_{2i-1}$.
2. Find $\sum_{i=1}^n (X_i + 1)$, $\sum_{i=1}^n (X_i - 10)$.
3. Find $\sum_{i=1}^n (10X_i)$, $\sum_{i=1}^n (X_i/10)$.
4. Find $\sum_{i=1}^n (-X_i)$, $\sum_{i=1}^n (-1)^i X_i$.

ALGEBRA PROBLEMS

1. Given three sets of n numbers each, prove
Theorem 1: $\Sigma(X_i + Y_i - Z_i) = \Sigma X_i + \Sigma Y_i - \Sigma Z_i$
2. Prove Theorem 2: $\Sigma(X_i + a) = \Sigma X_i + na$
3. Prove Theorem 3: $\Sigma cX_i = c\Sigma X_i$.
4. Prove $\Sigma(X_i - \bar{x}) = 0$

ALGEBRA PROBLEMS, (c't'd)

5. Prove $\frac{(\sum X_i)^2}{n} = n\bar{x}^2$

6. Prove $\sum (X_i - \bar{x})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$

7. Prove $\sum X_i (X_i - \bar{x}) = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$

8. Prove $\sum (X_i - a)^2 = \sum (X_i - \bar{x})^2 + n(\bar{x} - a)^2$

Does this entitle you to conclude that $\sum (X_i - \bar{x})^2$ is a minimum sum of squares? Explain.

Coding. The arithmetic which is a part of statistics is often simplified by coding the original data. For example, a set of numbers such as 1897.358, 1897.363, 1897.341, 1897.377, ... might be coded as 58, 63, 41, 77, The arithmetic operations such as computing a mean and a variance are carried out with the simpler set of numbers and the results are then decoded to apply to the original data.

Let X_1, X_2, \dots, X_n be the original data. Code as Y_1, Y_2, \dots, Y_n where $Y_i = a + bX_i$. Hence $X_i = (Y_i - a)/b$.

If we observe the coding carefully, the order of procedure is seen to be

- i. Multiply X_i by b
- ii. Add a to the result of step i.

Decoding is the reverse procedure involving inverse operations. We proceed to

- i. Subtract a from Y_i
- ii. Divide the result by b .

In computing a mean, coded values are used. We have

$$\begin{aligned}\bar{y} &= \frac{1}{n} \sum Y_i = \frac{1}{n} [\sum (a + bX_i)] \\ &= \frac{1}{n} [na + b\sum X_i] \text{ by Theorems 2 and 3} \\ &= \frac{na}{n} + b \frac{\sum X_i}{n} \\ &= a + b\bar{x}\end{aligned}$$

Hence $\bar{x} = (\bar{y} - a)/b$. Decoding a mean is seen to be the same operation as decoding an observation.

In computing a variance with coded values, only the step involving multiplication of a constant has any effect on the variance. (The numbers 1,2,3,4 are no more variable than the numbers 101,102,103,104.) Step ii of the coding process has no effect. Step i of the decoding process is not applied.

PROBLEMS

1. Prove that

$$s_Y^2 = a^2 s_X^2 \quad \text{and} \quad s_X^2 = \frac{s_Y^2}{a^2}$$

From the preceding problem, it follows that $s_Y = a s_X$ and $s_X = s_Y/a$.

2. Prove that for $n = 2$,

$$s^2 = \frac{(X_1 - X_2)^2}{2}$$

Simultaneous samples. It is far more common to have a sampling operation involve several populations rather than a single one. If a sampling operation involves two populations and the observations are paired, then it will be possible to consider that a single population, one of differences, has been sampled. This is the exception and the one-population approach is not necessary.

When two or more populations are sampled, it is possible to use different letters for the observations from the different populations. Thus we might use X_1, \dots, X_n and Y_1, \dots, Y_m to represent a sample of size n from a population of X 's and one of size m from a population of Y 's. However, a more common procedure is to use two subscripts, one to denote the sample and one for the observation within the sample.

Let X_{ij} represent the j -th observation on the i -th sample, $i = 1, \dots, k$ and $j = 1, \dots, n_i$. Thus we have k samples which are not necessarily of the same size since n_i need not equal n_j , and so on.

PROBLEMS OF NOTATION

Let 7,3,6,3; 3,2,7,9,5; 4,6,7,8,11; and 4,5,9,7,11,8 be samples obtained in a single experiment and from possibly different populations. Using the notation of the preceding paragraph, answer the following:

1. What is the value of k ? n_1 ? n_2 ? n_3 ? n_4 ?
2. What is the value of X_{11} ? X_{13} ? X_{22} ? X_{25} ? X_{33} ? X_{44} ? X_{55} ?
3. What is the value of X_{ij} for $i = 2, j = 4$? $i = 4, j = 1$?
 $i = 1, j = 5$?
4. What is the value of i when $X_{i2} = 3$? $X_{i3} = 7$? $X_{i4} = 11$?
5. What is the value of j when $X_{1j} = 3$? $X_{2j} = 7$? $X_{4j} = 8$?
6. What is the value of (i,j) when $X_{ij} = 7$? 4? 8?

Addition notation. Once again we use a Σ to denote summation. Now, however, we have to be especially careful about indices of summation. For example, $\sum_{j=1}^{n_i} X_{ij}$ says to add all the X_{ij} 's in the i -th sample. There will be k such quantities. Again, $\sum_{i=1}^k X_{ij}$ says to add the j -th observations from all samples. This is generally not very meaningful unless there is something common about the observations. For example, if we obtained a sample of n sets of twins, the symbol X_{ij} could be such that i referred to the individual in the set while j referred to the particular set. Then i would equal

1 or 2 and the 2's might receive a dietary supplement while the 1's would not. Here, $\sum_{i=1}^2 X_{ij}$ would be a total for the set of twins and there would be n such totals where n is the number of sets.

To avoid the use of summation signs and indices, a so-called dot notation is sometimes used. In this case, a dot replaces the summation sign and the index involved. Thus $\sum_{j=1}^{n_i} X_{ij} = X_{i.}$, $\sum_{i=1}^k X_{ij} = X_{.j}$, $\sum_{i,j}^{k,n_i} X_{ij} = X_{..}$ = the grand total.

PROBLEMS INVOLVING ADDITION

Using the numbers given for the preceding set of problems:

1. Find $\sum_j X_{ij}$ for $i = 1$. For $i = 4$.
2. Find $\sum_j X_{ij}$ for $j = 3$. For $j = 4$.
3. Find $X_{2.}$, $X_{4.}$, $X_{.1}$, $X_{.2}$.
4. Find $X_{.j}$ for $j = 3$.
5. How many values of $X_{i.}$ are there?
6. What is the value of $\bar{x}_{1.}$? $\bar{x}_{4.}$? $\bar{x}_{.1}$? $\bar{x}_{.3}$? $\bar{x}_{..}$?
7. Where a Σ appears in problems 1 and 2, replace by dot notation.
8. Where a dot appears in problems 3,4,5, replace by Σ notation.

PAIRED OBSERVATIONS: Suppose we have a sample of n pairs of observations. An observation is denoted by X_{ij} , $i = 1, 2$; $j = 1, \dots, n$. Set $X_{1j} - X_{2j} = D_j$. Then the mean is denoted by \bar{d} and $d_j = D_j - \bar{d}$ is a deviate.

PROBLEMS

1. Show that $\bar{d} = \bar{x}_{1.} - \bar{x}_{2.}$.
2. Show that $\sum_j d_j^2 = \sum x_{1j}^2 + \sum x_{2j}^2 - 2\sum x_{1j}x_{2j}$.

Simultaneous samples - unpaired observations. When more than one sample is taken, an observation may be denoted by X_{ij} where i refers to the sample and j to the observation within the sample. When the observations are unpaired, the j -th observation in the first sample is no more like the j -th observation in the second sample than it is like any other. Thus a $\sum_i X_{ij}$ is not meaningful. This is not so in the case of paired samples where the j -th observations might be twins. Here a $\sum_i X_{ij}$ is meaningful.

Consider the simple case of two samples, $i = 1, 2$, in the general situation where the samples are of unequal size, n_1 and n_2 .

To compute a variance to be used as a yardstick in answering the usual statistical questions it is necessary to eliminate the effect of the difference between means. Thus if the samples are from populations with very different means but a common variance, the overall variance would be much larger than that within either sample. It is customary to pool the two sample variances in a way which weights each according to the number of d.f. associated with it; the larger the sample the greater the weight. The result is called a "within sample" variance or a pooled variance. It is calculated as

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

PROBLEM: Show that

$$s_p^2 = \frac{\sum_j X_{1j}^2 - \frac{(\sum_j X_{1j})^2}{n_1} + \sum_j X_{2j}^2 - \frac{(\sum_j X_{2j})^2}{n_2}}{(n_1-1) + (n_2-1)}$$

i.e. the pooled sum of squares divided by the pooled d.f.

It is clear that the mean $\bar{x}_{..} = \frac{\sum_{i,j} X_{ij}}{n_1+n_2}$, is a weighted average of the sample means.

PROBLEM: Show that

$$\bar{x}_{..} = \frac{n_1 \bar{x}_{1.} + n_2 \bar{x}_{2.}}{n_1 + n_2}$$

i.e. the weighted sum of the means divided by the sum of the weights.

When a variance among means is required, it is customary to compute a weighted variance.

PROBLEM: Show that the weighted sum of squares

$$\begin{aligned}\sum n_i (\bar{x}_{i.} - \bar{x}_{..})^2 &= n_1 \bar{x}_{1.}^2 + n_2 \bar{x}_{2.}^2 - (n_1 + n_2) \bar{x}_{..}^2 \\ &= \frac{(\sum x_{1j})^2}{n_1} + \frac{(\sum x_{2j})^2}{n_2} - \frac{(\sum x_{ij})^2}{n_1 + n_2} \\ &= (\bar{x}_1 - \bar{x}_2)^2 \frac{n_1 n_2}{n_1 + n_2}\end{aligned}$$

It is also to be noted that the total sum of squares is the sum of the within-sample sum of squares and the weighted sum of squares among the means. This may be written as in the following problem.

PROBLEM: Show that

$$(n_1 + n_2 - 1)s^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2.$$

This property is often used as a method of obtaining the error sum of squares. From the total sum of squares, subtract the sum of squares attributable to means.

The test criterion for the hypothesis of no difference between the means of the populations sampled is Student's t or Snedecor's F.

PROBLEM: Show that for unpaired samples of equal size

$$\begin{aligned}t^2 &= \left[\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{2}{n} \right)}} \right]^2 \\ &= \frac{n \left[\sum_{i=1}^2 \bar{x}_{i.}^2 - \frac{(\sum \bar{x}_{i.})^2}{2} \right]}{s^2} = F\end{aligned}$$

Note that the numerator of F is a multiple of the variance of means.

Simultaneous samples - the completely random design or one-way classification

In the case of a well-planned and executed experiment, the arithmetic of the analysis of variance will be simple. For the usual designs, this means that the various directly-computed sums of squares are additive. For example, a treatment sum of squares may be computed directly from treatment totals or means; error sums of squares are found by subtracting the directly-computed sums of squares from a total sum of squares. Direct computation is permissible because of independence of effects as represented by any two lines in the analysis of variance.

Algebraically, independence implies that certain cross products are zero. To see this, consider the simple case of the completely random design. The model is given by

$$X_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, k, j = 1, \dots, M_i$$

The parameter μ is estimated by \bar{x} and τ_i by $\bar{x}_{i.} - \bar{x}$, estimates such that the residual sum of squares is a minimum. In terms of the estimates of the parameters, we may write an observation as

$$X_{ij} = \bar{x} + (\bar{x}_{i.} - \bar{x}) + (X_{ij} - \bar{x}_{i.})$$

Consequently

$$X_{ij} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (X_{ij} - \bar{x}_{i.})$$

and

$$\sum_{i,j} (X_{ij} - \bar{x})^2 = \sum_{i,j} (\bar{x}_{i.} - \bar{x})^2 + 2 \sum_{i,j} (\bar{x}_{i.} - \bar{x})(X_{ij} - \bar{x}_{i.}) + \sum_{i,j} (X_{ij} - \bar{x}_{i.})^2$$

PROBLEMS:

$$1. \text{ Show } \sum_{i,j} (\bar{x}_{i.} - \bar{x})(X_{ij} - \bar{x}_{i.}) = 0$$

$$2. \text{ Show } \sum_{i,j} (\bar{x}_{i.} - \bar{x})^2 = n \sum_i (\bar{x}_{i.} - \bar{x})^2 = \frac{\sum_i (Y_{i.})^2}{n} - C \text{ where } C = \frac{(\sum_{i,j} Y_{ij})^2}{kn}$$

for n observations per treatment.

3. Rewrite problems 1 and 2 for the case of unequal sample sizes and complete the proofs.

The randomized complete block design. In this case, we have a true two-way classification. Here both subscripts refer to a classification whereas the second was only a tag for easy reference to an observation in the case of the completely random design. The model is

$$X_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}$$

where $i = 1, \dots, b$ blocks and $j = 1, \dots, t$ treatments. It is fairly obvious that μ is estimated by \bar{x} , ρ_i by $\bar{x}_{i.} - \bar{x}$, and τ_j by $\bar{x}_{.j} - \bar{x}$. We may write

$$X_{ij} = \bar{x} + (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})$$

or

$$X_{ij} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}).$$

It follows that

$$\begin{aligned} \sum_{i,j} (X_{ij} - \bar{x})^2 &= \sum_{i,j} (\bar{x}_{i.} - \bar{x})^2 + 2 \sum_{i,j} (\bar{x}_{i.} - \bar{x})(\bar{x}_{.j} - \bar{x}) + 2 \sum_{i,j} (\bar{x}_{i.} - \bar{x})(X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) \\ &+ \sum_{i,j} (\bar{x}_{.j} - \bar{x})^2 + 2 \sum_{i,j} (\bar{x}_{.j} - \bar{x})(X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) + \sum_{i,j} (X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \end{aligned}$$

PROBLEMS:

1. Show that the three cross-product terms are separately equal to zero.
2. Show $\sum_{i,j} (\bar{x}_{i.} - \bar{x})^2 = t \sum_i (\bar{x}_{i.} - \bar{x})^2$

$$= \frac{\sum_i X_{i.}^2}{t} - C$$

$$\text{where } C = \frac{(\sum_{i,j} X_{ij})^2}{tb}$$

The Latin Square. In the Latin square, we have a three-way classification where two subscripts are actually sufficient to locate an observation, the third tells what treatment was received by the observation. Consider a square of side k . We may write the model as

$$X_{ij}(t) = \mu + \rho_i + K_j + \tau(t) + \epsilon_{ij}.$$

The subscript t is in brackets to emphasize the fact that there are k^2 , not k^3 , observations; any t is a set of i, j values. Estimate μ by \bar{x} , ρ_i by $\bar{x}_{i.} - \bar{x}$, K_j by $\bar{x}_{.j} - \bar{x}$, and $\tau(t)$ by $\bar{x}(t) - \bar{x}$ where $\bar{x}(t)$ refers to a treatment mean. Again,

$$X_{ij}(t) = \bar{x} + (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (\bar{x}(t) - \bar{x}) + (X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} - \bar{x}(t) + 2\bar{x})$$

$$\text{or } X_{ij}(t) - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (\bar{x}(t) - \bar{x}) + (X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} - \bar{x}(t) + 2\bar{x})$$

Expected Values: A term commonly met in statistical literature is "expected value". It is nothing more than a population average. Thus, if X is a variable from a population with mean μ , then the expected value of X is μ . Again, if we compute all possible values of $s^2 = \frac{\sum (X_i - \bar{x})^2}{n-1}$ for fixed n , a population of s^2 's is obtained. The statement that this s^2 is unbiased is equivalent to the statement that the average value or expected value of s^2 is σ^2 .

The letter E is commonly used to indicate an expected value or the process of finding an expectation. Thus from the above paragraph $E(X) = \mu$ and $E(s^2) = \sigma^2$. Only the second statement requires proof.

Calculation of Expected Values: For a finite population, it is easy to calculate expected values. Suppose we are dealing with an unbiased die. The probability distribution involved is given by $P(X = X_i) = 1/6$, $X_i = 1, 2, \dots, 6$. I.e., the probability that the random variable X takes on the value X_i (which may be 1, 2, ..., or 6) is $1/6$. To find the expected value of X for this distribution, each value must be weighted by the probability with which it occurs, i.e.,

$$\begin{aligned}\mu = E(X) &= \sum_{i=1}^6 X_i P(X = X_i) \\ &= 1(1/6) + 2(1/6) + 3(1/6) + \dots + 6(1/6) \\ &= 3 \frac{1}{2}\end{aligned}$$

Problem: Find $E(X^2)$ for the above distribution.

In general, for a finite population, one finds an expected value by adding all possible values and dividing by the total number. Where a number occurs more than once, it must be added as often as it occurs. This will generally mean multiplying it by its frequency before adding. Or, all numbers may be multiplied by their relative frequencies in which case the divisor (the total number) is not required. The sum of the relative frequencies is unity, i.e., they are probabilities.

Where an infinite population, such as the normal, is involved, the definition is obviously not applicable. The arithmetic becomes a little fancier, new symbols and terms are required. Actually, with a few rules which are fairly obvious, simple algebra is sufficient for solving many problems.

Rules of Operation with Expected Values: The letter E is sometimes referred to as an operator. It indicates that a certain operation is required, the operation of taking an expectation. These rules are:

1. $E(k) = k$ where k is a constant.

This is obvious from the definition of an expected value.

$E(k) = \sum_i kP(X = X_i) = k \sum_i P(X = X_i) = k \cdot 1$ since the sum of all the probabilities is unity.

2. $E(kX) = kE(X)$.

This is again obvious from the definition.

3. $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$.

Here, the X_i 's refer to different variables rather than different values of a single variable. For example, this might be the case of tossing n dice. This says that the average of a sum equals the sum of the averages.

- 3a. $E(k_1X_1 + k_2X_2 + \dots + k_nX_n) = k_1E(X_1) + k_2E(X_2) + \dots + k_nE(X_n)$.

This is a combination of rules 2 and 3.

Problem: (Use the rules): What is the expected value of the sample mean, \bar{x} ?

By definition, $\sigma^2 = E[(X - \mu)^2]$. I.e., the variance is the average value of $(X - \mu)^2$.

Problem: Show that $\sigma^2 = E(X^2) - \mu^2$.

Independence in Random Sampling: A sample of one observation can be plotted as a point on a line. A sample of two requires a plane unless one throws away some of the information, as in computing a mean. A probability distribution for a sample of one observation is shown as a series of lines vertical to the axis (discrete variable) or as a curve above the axis (continuous variable). For a sample of two, a series of lines sticking up through a plane, or a surface (e.g., a hemisphere) above a plane is required. Distributions for sets of observations, such as a sample, are called joint probability distributions.

In random sampling of a continuous variable or of a discrete one with replacement,

$$E(X_i X_j) = E(X_i)E(X_j)$$

Subscripts indicate the order of sampling. I.e., X_i is the i -th observation drawn.

Problem: Show that $E(X_i - \mu)(X_j - \mu) = 0$ in random sampling as indicated above.

VARIANCE OF A MEAN. By definition

$$\begin{aligned}
 \sigma_{\bar{x}}^2 &= E(\bar{x} - \mu)^2 \\
 &= E\left(\frac{\sum X_i}{n} - \frac{n\mu}{n}\right)^2 \\
 &= E\left(\frac{\sum (X_i - \mu)}{n}\right)^2 \\
 &= \frac{1}{n^2} E[(X_1 - \mu)^2 + \dots + (X_n - \mu)^2 + 2(X_1 - \mu)(X_2 - \mu) \\
 &\quad + 2(X_1 - \mu)(X_3 - \mu) + \dots + 2(X_{n-1} - \mu)(X_n - \mu)] \\
 &= \frac{1}{n^2} E[\sigma^2 + \dots + \sigma^2 + 2.0 + 2.0 + \dots + 2.0] \\
 &\quad \text{(By definition and previous problem)} \\
 &= \frac{1}{n^2} \cdot n\sigma^2 \\
 &= \sigma^2/n
 \end{aligned}$$

UNBIASEDNESS OF s^2 . Consider the numerator of s^2 , i.e. $\sum (X_i - \bar{x})^2$.

$$\begin{aligned}
 E[\sum (X_i - \bar{x})^2] &= E\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right] \\
 &= \sum E(X_i^2) - E\left[\frac{(\sum X_i)^2}{n}\right]
 \end{aligned}$$

(Note one term is the sum of expected values squared while the other is the expected value of a squared sum. The latter will involve cross-products.)

$$= \sum E(X_i^2) - \frac{1}{n} E[\sum X_i^2 + 2X_1X_2 + 2X_1X_3 + \dots + 2X_{n-1}X_n]$$

$$= \sum E(X_i^2) - \frac{1}{n} \sum E(X_i^2) - \frac{2}{n} \frac{n(n-1)}{2}$$

$$= n[\sigma^2 + \mu^2] - \frac{n}{n} [\sigma^2 + \mu^2] - (n-1)\mu \cdot \mu$$

(See problem in section on rules.)

$$= (n-1)\sigma^2 + (n-1)\mu^2 - (n-1)\mu^2$$

$$= (n-1)\sigma^2.$$

$$\therefore E(s^2) = \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2.$$

ALGEBRA NOTES
ON
REGRESSION AND CORRELATION

Given a population of (X, Y) 's, the regression of Y on X is defined as the line traced by the mean values of Y . Each mean value is that for a population of Y 's having a specified X -value.

Where the regression of Y on X is linear, we may write

$$\mu_{Y.X} = \alpha + \beta X.$$

Here, α and β are numerical constants and we need only substitute the desired value of X in order to find the numerical value of the mean of the Y 's having that value of X . For a sample, the regression of Y on X is usually written as

$$\begin{aligned} Y &= a + bX \\ &= \bar{y} + b(X - \bar{x}) \end{aligned}$$

and the notion of a population mean or its estimate is obscured. This may be remedied to some extent by placing a $\hat{}$ over the Y to give \hat{Y} . Here, a and b are estimates of α and β .

In estimating any population mean, the usual procedure is to calculate an estimate such that the sum of the squares of the differences between the observed value and the estimate is a minimum. In the process, the sum of the differences is found to be zero.

It can be shown that

$$\Sigma(Y - \hat{Y}) = 0$$

and

$$\Sigma(Y - \hat{Y})^2 = \text{a minimum}$$

when Y is an observed value and \hat{Y} is the corresponding estimate of $\mu_{Y.X}$ as obtained from the previous equation.

Problem 1: Show $\Sigma(Y - \hat{Y}) = 0$.

Problem 2: Show that $\Sigma(Y - \hat{Y})^2 = \Sigma(Y - \bar{y})^2 - b^2 \Sigma(X - \bar{x})^2$

(Hint: Replace \hat{Y} by $\bar{y} + b(X - \bar{x})$ and use the definition of b .)

To show that $\Sigma(Y - \hat{Y})^2$ is a minimum, replace \hat{Y} by $\tilde{Y} = \bar{y} + (b+k_1)(X-\bar{x}) + k_2$,

essentially a new line with a different origin and slope.

Consider

$$Y - \tilde{Y} = (Y - \bar{y}) - (b + k_1)(X - \bar{x}) - k_2.$$

Square this quantity as indicated by the brackets

$[(a - b - c)^2 = a^2 + b^2 + c^2 - 2ab - 2ac + 2bc]$ and consider the sum of squares, $\Sigma(Y - \tilde{Y})^2$. This can now be shown to be equal to the answer of problem 2 plus terms which are either zero or positive.

Problem 3: Complete the proof just outlined.

Since the Y's are required to be randomly drawn from the populations specified by the X-values, the variance of $b = \Sigma x_i y_i / \Sigma x_i^2$ is easily calculated.

In general, the variance of the linear function $\Sigma a_i y_i$ is given by

$\Sigma a_i^2 \sigma_i^2 + 2 \Sigma_{i < j} a_i a_j \sigma_{ij}$. For b, the y's are independent so that the second term is zero.

Problem 4: For b, what is a_i ? Show that the variance of b is $\sigma_{Y \cdot X}^2 / \Sigma x_i^2$.

The portion of the sum of squares of Y which can be attributed to variation in the X's, is given by $(\Sigma xy)^2 / \Sigma x^2$.

Problem 5: Show that $(\Sigma xy)^2 / \Sigma x^2 = b^2 \Sigma x^2$.

The correlation coefficient is a useful and meaningful quantity when the pairs of observations are randomly drawn. The square of r is the fraction of the total sum of squares for the dependent variable which can be attributed to variation in the independent variable.

Problem 6: Show that $(\Sigma xy)^2 / \Sigma x^2 = r^2 \Sigma y^2$.

When two regressions are meaningful, either variable may be regarded as the independent one. The regressions of Y on X and of X on Y are denoted by $b_{Y \cdot X}$ and $b_{X \cdot Y}$ respectively.

Problem 7: Show that $r^2 = b_{Y \cdot X} b_{X \cdot Y}$.

Problem 8: Show that $b_{Y \cdot X} = r \frac{s_y}{s_x}$.

In testing b for significance, t may be used as the criterion,

$$t = \frac{b}{s_{Y \cdot X} / \sqrt{\Sigma x^2}} \quad \text{or } F \text{ may be used, } F = \frac{\text{Reduction in } \Sigma y^2}{(\text{Reduced } \Sigma y^2) / (n - 2)}.$$

Problem 9: Show that $t^2 = F$.

Derivation and solution of certain linear equations
arising in statistics

Although the general class-room approach (Plant Breeding 210 and 211) to the problem of testing hypotheses may seem to imply that a criterion (t , χ^2 , F , etc.) is decided upon and then justified, there are general theories which give rise to test criteria. One of these is the so-called maximum likelihood theory which gives rise to the test criteria so far discussed as well as others. This theory results in minimizing the residual sum of squares when the data are assumed to have a normal distribution. Thus, we could equally well talk about a least-squares procedure in these cases.

Least squares equations can be derived by following a simple set of rules without any special knowledge of the mathematics that gives rise to them.

Consider the simple case described by the model

$$X_i = \mu + \epsilon_i \quad \text{or} \quad X_i - \mu = \epsilon_i$$

The problem is to estimate μ , say by $\hat{\mu}$, so that

$$1. \quad \Sigma(\text{residual})^2 = \Sigma(X_i - \hat{\mu})^2 = \text{minimum}$$

The estimating equation is given by

$$2. \quad -2\Sigma(X_i - \hat{\mu}) = 0.$$

$$\therefore \quad \Sigma X_i - n\hat{\mu} = 0 \quad \text{or} \quad \hat{\mu} = \frac{\Sigma X_i}{n} = \bar{x}.$$

In equation 2, the 2 comes from the power 2 in the sum of squares. Since it is present in every least squares equation, it may be ignored entirely. The -1 is the coefficient of $\hat{\mu}$ in the equation. One cannot generally ignore this coefficient though the sign may be ignored since it affects the total equation rather than a part of it. The other property to be noted is that equation 2 involves rewriting equation 1 without the square.

Consider the case where two parameters are to be estimated to give a regression equation. The model is

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

and parameters α and β are to be estimated by $\hat{\alpha}$ and $\hat{\beta}$, say. The problem is to find $\hat{\alpha}$ and $\hat{\beta}$ such that

$$\Sigma(\text{residual})^2 = \Sigma(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 = \text{minimum}.$$

The estimating equations are given by

$$\begin{aligned} 3. \quad \Sigma(Y_i - \hat{\alpha} - \hat{\beta}X_i) &= 0, \quad \text{the } \hat{\alpha} \text{ equation,} \\ \Sigma X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i) &= 0, \quad \text{the } \hat{\beta} \text{ equation.} \end{aligned}$$

In generalizing about least squares equations, note that we require a re-writing of the least squares equation without the square, that the coefficient of the estimator in the equation becomes a multiplier of the equation, that this multiplier is properly inside the summation sign but that its sign may be ignored, and that there are as many equations as parameters to be estimated. These rules will require some modification when certain subscripts are required on the parameters but, meanwhile, may be summarized as follows:

- i. Write, in brackets, the expression for the general residual as often as there are parameters to be estimated.
- ii. Multiply the expressions successively by the coefficients of the parameters to be estimated.
- iii. Precede each of the expressions in ii. by a summation sign and equate to zero.

Let us now solve equations 3. They become

$$\Sigma Y_i - n\hat{\alpha} - \hat{\beta}\Sigma X_i = 0$$

from which

$$\begin{aligned} \hat{\alpha} &= \frac{\Sigma Y_i}{n} - \hat{\beta} \frac{\Sigma X_i}{n} \\ &= \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} \text{ not yet available,} \end{aligned}$$

and

$$\Sigma X_i Y_i - \hat{\alpha}\Sigma X_i - \hat{\beta}\Sigma X_i^2 = 0$$

or

$$\begin{aligned} \hat{\beta}\Sigma X_i^2 &= \Sigma X_i Y_i - \hat{\alpha}\Sigma X_i \\ &= \Sigma X_i Y_i - (\bar{y} - \hat{\beta}\bar{x})\Sigma X_i, \text{ using the } \hat{\alpha} \text{ equation,} \end{aligned}$$

or

$$\hat{\beta}(\Sigma X_i^2 - \bar{x}\Sigma X_i) = \Sigma X_i Y_i - \bar{y}\Sigma X_i$$

from which

$$\hat{\beta} = \frac{\Sigma X_i Y_i - \frac{(\Sigma Y_i)(\Sigma X_i)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}} = \frac{\Sigma xy}{\Sigma x^2}$$

Proceeding to the general k-variate regression problem, we wish to minimize

$$\Sigma(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})^2.$$

Following the three rules, the $k+1$ equations become

$$\Sigma(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki}) = 0, \text{ the } \hat{\alpha} \text{ equation}$$

$$\Sigma X_{1i}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki}) = 0, \text{ the } \hat{\beta}_1 \text{ equation}$$

...

$$\Sigma X_{ki}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki}) = 0, \text{ the } \hat{\beta}_k \text{ equation.}$$

Problem 1: Obtain the three equations for $k = 2$ and solve.

The Expected Value and Variance of the Sample Regression Coefficient

By definition, the regression of Y on X is given by the equation

$$\mu_{Y \cdot X} = E(Y_X) = \alpha + \beta X,$$

i.e. for each value of X there is a mean Y, $\mu_{Y \cdot X}$, dependent upon X. This also states that there is a distribution of Y-values for each X and that X appears as a parameter in any distribution of Y.

The least-squares procedure leads to estimates of α and β , namely

$$\begin{aligned}\hat{\alpha} &= a = \bar{y} - b\bar{x} \\ \hat{\beta} &= b = (\Sigma xy) / \Sigma x^2.\end{aligned}$$

If the estimates are such that $E(a) = \alpha$ and $E(b) = \beta$, then the estimates are said to be (mean-) unbiased estimates of α and β .

Consider $E(b)$.

$$\begin{aligned}E(b) &= E\left(\frac{\Sigma xy}{\Sigma x^2}\right) \\ &= \frac{1}{\Sigma x^2} [\Sigma x_i E(Y_i - \bar{y})] \\ &= \frac{1}{\Sigma x^2} [\Sigma x_i E(Y_i) - \Sigma x_i E(\bar{y})] \\ &= \frac{1}{\Sigma x^2} [\Sigma x_i (\alpha + \beta X_i)] , \text{ by definition and since } \Sigma x_i = 0 \\ &= \frac{1}{\Sigma x^2} [\alpha \Sigma x_i + \beta \Sigma x_i X_i] \\ &= \beta\end{aligned}$$

Problem 1: Show $\Sigma x_i X_i = \Sigma x_i^2$.

$$\begin{aligned}E(a) &= E(\bar{y} - b\bar{x}) \\ &= E(\bar{y}) - \bar{x}E(b) \quad \text{since } \bar{x} \text{ is a parameter of the distribution} \\ &= \mu_{Y \cdot \bar{x}} - \bar{x}\beta \\ &= \alpha\end{aligned}$$

Consider the variance of b. By definition,

$$\begin{aligned}\sigma_b^2 &= E(b - E(b))^2 \\ &= E(b^2) - \beta^2\end{aligned}$$

We will use the following relation. Since

$$\sigma_{Y \cdot X}^2 = E(Y_i)^2 - [E(Y_i)]^2, \text{ constant for all } X,$$

Then

$$E(Y_i)^2 = \sigma_{Y \cdot X}^2 + E^2(Y_i).$$

Now

$$\begin{aligned}E(b^2) &= E\left(\frac{\sum xy}{(\sum x^2)^2}\right)^2 = E\left(\frac{\sum xY}{(\sum x^2)^2}\right)^2 \\ &= E\left(\frac{\sum_i x_i^2 Y_i^2}{(\sum x^2)^2}\right) + E\left(\frac{\sum_{i \neq j} x_i x_j Y_i Y_j}{(\sum x^2)^2}\right) \\ &= \frac{\sum_i x_i^2 E(Y_i^2)}{(\sum x^2)^2} + \frac{\sum_{i \neq j} x_i x_j E(Y_i) E(Y_j)}{(\sum x^2)^2} \\ &= \frac{\sum_i x_i^2 (\sigma_{Y \cdot X}^2 + \mu_{Y \cdot X_i}^2)}{(\sum x^2)^2} + \frac{\sum_{i \neq j} x_i x_j \mu_{Y \cdot X_i} \mu_{Y \cdot X_j}}{(\sum x^2)^2} \\ &= \frac{\sigma_{Y \cdot X}^2}{\sum x^2} + \frac{\sum_i x_i^2 \mu_{Y \cdot X_i}^2}{(\sum x^2)^2} + \frac{\sum_{i \neq j} x_i x_j \mu_{Y \cdot X_i} \mu_{Y \cdot X_j}}{(\sum x^2)^2} \\ &= \frac{\sigma_{Y \cdot X}^2}{\sum x^2} + \frac{(\sum_i x_i \mu_{Y \cdot X_i})^2}{(\sum x^2)^2} \\ &= \frac{\sigma_{Y \cdot X}^2}{\sum x^2} + \frac{[\sum_i x_i (\mu_{Y \cdot X} + \beta(x_i - \bar{x}))]^2}{(\sum x^2)^2} \\ &= \frac{\sigma_{Y \cdot X}^2}{\sum x^2} + \frac{[\mu_{Y \cdot X} \sum x_i + \beta \sum x_i (x_i - \bar{x})]^2}{(\sum x^2)^2} \\ &= \frac{\sigma_{Y \cdot X}^2}{\sum x^2} + \beta^2\end{aligned}$$

$$\sigma_b^2 = E(b^2) - \beta^2$$

$$= \frac{\sigma_{Y \cdot X}^2}{\sum x^2}$$

By definition, the variance of a is given by

$$\begin{aligned}\sigma_a &= E(a^2) - E^2(a) \\ E(a^2) &= E(\bar{y} - b\bar{x})^2 \\ &= E(\bar{y}^2) - 2\bar{x}E(b\bar{y}) + \bar{x}^2E(b^2)\end{aligned}$$

First,

$$\begin{aligned}E(\bar{y}^2) &= \frac{\sigma_{Y \cdot X}^2}{n} + E^2(\bar{y}) \\ &= \frac{\sigma_{Y \cdot X}^2}{n} + \mu_{Y \cdot \bar{x}}^2\end{aligned}$$

Second,

$$-2\bar{x}E(b\bar{y}) = -\frac{2\bar{x}}{\Sigma x_i^2} \Sigma x_i E(Y_i \bar{y})$$

and for fixed i ,

$$\begin{aligned}E(Y_i \bar{y}) &= E(Y_i \frac{\Sigma Y_j}{n}) \\ &= \frac{1}{n} \left\{ E(Y_i^2) + E(Y_i \sum_{j \neq i} Y_j) \right\} \\ &= \frac{1}{n} \left\{ \sigma_{Y \cdot X}^2 + \mu_{Y \cdot X_i}^2 + \mu_{Y \cdot X_i} \sum_{j \neq i} \mu_{Y \cdot X_j} \right\} \\ &= \frac{1}{n} \left\{ \sigma_{Y \cdot X}^2 + \mu_{Y \cdot X_i} \sum_j \mu_{Y \cdot X_j} \right\} \\ &= \frac{1}{n} \left\{ \sigma_{Y \cdot X}^2 + n \mu_{Y \cdot X_i} \mu_{Y \cdot \bar{x}} \right\}\end{aligned}$$

Problem: Show $\sum_i \mu_{Y \cdot X_i} = n \mu_{Y \cdot \bar{x}}$

Hence,

$$\begin{aligned}-2\bar{x}E(b\bar{y}) &= -\frac{2\bar{x}}{\Sigma x_i^2} \Sigma x_i \left(\frac{\sigma_{Y \cdot X}^2}{n} + \mu_{Y \cdot X_i} \mu_{Y \cdot \bar{x}} \right) \\ &= -\frac{2\bar{x}}{\Sigma x_i^2} (\mu_{Y \cdot \bar{x}} \Sigma x_i \mu_{Y \cdot X_i}) \\ &= -\frac{2\bar{x}}{\Sigma x_i^2} \left\{ \mu_{Y \cdot \bar{x}} \left[\Sigma x_i \mu_{Y \cdot \bar{x}} + \beta \Sigma x_i (X_i - \bar{x}) \right] \right\} \\ &= -\frac{2\bar{x}}{\Sigma x_i^2} \mu_{Y \cdot \bar{x}} \beta \Sigma x_i^2 \\ &= -2\bar{x} \mu_{Y \cdot \bar{x}} \beta\end{aligned}$$

$$\text{Finally } \bar{x}^2 E(b^2) = \bar{x}^2 \left(\frac{\sigma_{Y \cdot X}^2}{\Sigma x^2} + \beta^2 \right)$$

$$\begin{aligned} \text{Hence } E(a^2) - E^2(a) &= \frac{\sigma_{Y \cdot X}^2}{n} + \mu_{Y \cdot \bar{x}}^2 - 2\bar{x}\mu_{Y \cdot \bar{x}}\beta + \frac{\bar{x}^2 \sigma_{Y \cdot X}^2}{\Sigma x^2} + \bar{x}^2 \beta^2 \\ &\quad - \mu_{Y \cdot \bar{x}}^2 + 2\mu_{Y \cdot \bar{x}}\beta\bar{x} - \beta^2 \bar{x}^2 \\ &= \sigma_{Y \cdot X}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma x^2} \right) \\ &= \sigma_{Y \cdot X}^2 \left(\frac{\Sigma x^2}{n \Sigma x^2} \right) \end{aligned}$$

By definition, the covariance of a and b is given by

$$E(ab) - E(a)E(b)$$

$$\begin{aligned} E(ab) &= E(\bar{y} - b\bar{x})b \\ &= E(b\bar{y}) - \bar{x}E(b^2) \\ &= \mu_{Y \cdot \bar{x}}\beta - \bar{x} \left(\frac{\sigma_{Y \cdot X}^2}{\Sigma x^2} + \beta^2 \right) \\ &= (\mu_{Y \cdot \bar{x}} - \bar{x}\beta)\beta - \bar{x} \frac{\sigma_{Y \cdot X}^2}{\Sigma x^2} \\ &= \alpha\beta - \bar{x} \frac{\sigma_{Y \cdot X}^2}{\Sigma x^2} \end{aligned}$$

$$\begin{aligned} \text{Cov}(ab) &= \alpha\beta - \bar{x} \frac{\sigma_{Y \cdot X}^2}{\Sigma x^2} - \alpha\beta \\ &= \frac{-\bar{x} \sigma_{Y \cdot X}^2}{\Sigma x^2} \end{aligned}$$

Solutions of Linear Equations

Consider the equation

$$a_1X + b_1Y + c_1 = 0. \quad (1)$$

Provided at least one of the constants a_1 and b_1 is not zero, this equation can be interpreted as a straight line in the (X,Y) -plane. This line is the locus of all points (X_1, Y_1) which satisfy the equation. Any pair of values which satisfies the equation is said to be a solution of the equation.

Now if we have a second such equation

$$a_2X + b_2Y + c_2 = 0, \quad (2)$$

we have a second line. If these two lines intersect, i.e. if they are not parallel, the coordinates of the point of intersection satisfy both equations. We have a common solution for equations (1) and (2); or, we have solved the simultaneous equations.

Now if we have a third equation

$$a_3X + b_3Y + c_3 = 0, \quad (3)$$

the corresponding line may pass through the point of intersection of the other pair but need not.

In case three equations have a common solution, i.e. the three lines have a common point of intersection, any two, generally, of the equations are sufficient to obtain a solution of the set of equations; the third equation turns out to be a linear function of the other two, i.e. $m(1) + n(2) = (3)$ where m and n are constants and (1), (2) and (3) are the equations. For example, consider the equations

$$X = 2 \quad (4)$$

$$Y = 3 \quad (5)$$

$$X + 2Y = 8. \quad (6)$$

It is readily seen that $1(4) + 2(5) = 6$. It is also seen that any two equations are capable of supplying a solution but that the solution actually consists of equations (4) and (5). Had the last equation been $X + 2Y = 6$, then there would be no solution common to the three equations.

Consider the equations

$$X + Y = 2 \quad \text{and} \quad X + Y = 6.$$

Here are two equations, but they do not have a common solution. Graphically, they give parallel lines.

In general, n equations in n unknowns are necessary and sufficient to obtain a common solution. However, we have seen that certain conditions must be imposed upon them in order that they can be solved.

Two equations in two unknowns. Consider the two general equations (1) and (2) where at least one of c_1 and c_2 is not zero. The obvious method of solution is to solve one, say (1), for X in terms of Y and substitute this in the other. We obtain

$$\begin{aligned} X &= \frac{-b_1 Y - c_1}{a_1} \\ a_2 \left(\frac{-b_1 Y - c_1}{a_1} \right) + b_2 Y + c_2 &= 0 \\ &= \left(b_2 - \frac{a_2 b_1}{a_1} \right) Y + \left(c_2 - \frac{a_2 c_1}{a_1} \right) = 0 \end{aligned}$$

$$\begin{aligned} \text{or } Y &= \frac{a_2 c_1 - a_1 c_2}{a_1} \cdot \frac{a_1}{a_1 b_2 - a_2 b_1} \\ &= \frac{a_2 c_1 - a_1 c_2}{a_1 b_2 - a_2 b_1} \end{aligned}$$

$$\begin{aligned} \text{and } X &= \frac{-b_1 \left(\frac{a_2 c_1 - a_1 c_2}{a_1 b_2 - a_2 b_1} \right) - c_1}{a_1} \\ &= \frac{-a_2 b_1 c_1 + a_1 b_1 c_2 - a_1 b_2 c_1 + a_2 b_1 c_1}{a_1 (a_1 b_2 - a_2 b_1)} \\ &= \frac{b_1 c_2 - b_2 c_1}{a_1 b_2 - a_2 b_1} \end{aligned}$$

Here the denominators are the same while the numerators differ in that a_1 replaces b_1 and the sign changes. Obviously, we have no meaningful solution for our equations if $a_1 b_2 - a_2 b_1 = 0$.

The result $a_1 b_2 - a_2 b_1 = 0$ is sometimes written as

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \quad (7)$$

It is called a determinant. We may now rewrite our solution of equations (1) and (2) as

$$X = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \quad Y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}$$

Problems:

1. Find the numerical values of

$$\begin{vmatrix} 2 & 3 \\ 5 & 7 \end{vmatrix}, \quad \begin{vmatrix} 1 & -9 \\ 3 & 3 \end{vmatrix}, \quad \begin{vmatrix} 3 & -2 \\ -2 & 4 \end{vmatrix}$$

$$\begin{vmatrix} 2+3 & 3 \\ 5+7 & 7 \end{vmatrix}, \quad \begin{vmatrix} -1 & -9 \\ -3 & 3 \end{vmatrix}, \quad \begin{vmatrix} -2 & 3 \\ 4 & -2 \end{vmatrix}$$

$$\begin{vmatrix} 2 & 5 \\ 3 & 7 \end{vmatrix}, \quad \begin{vmatrix} 1 & -9 \\ 1+2(3) & -9+2(3) \end{vmatrix}, \quad \begin{vmatrix} 3 & -2 \\ 2(-2) & 2(4) \end{vmatrix}.$$

2. Do the previous numerical results suggest that we look for certain generalizations? Elaborate!

3. Show that the solution, obtained by the method of determinants, for equations

$$\hat{\alpha} + \bar{x}\hat{\beta} = \bar{y}$$

$$(\sum X_i)\hat{\alpha} + (\sum X_i^2)\hat{\beta} = \sum X_i Y_i$$

is $\hat{\alpha} = \bar{y} + \frac{\sum X_i Y_i}{\sum X_i^2} \bar{x}$ and $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}.$

4. Let $x' = \frac{X - \bar{X}}{s_x}$ and $y' = \frac{Y - \bar{Y}}{s_y}$. Show that the regression equations of X on Y and Y on X are given by

$$\begin{vmatrix} x' & y' \\ r & 1 \end{vmatrix} = 0 \quad \text{and} \quad \begin{vmatrix} 1 & r \\ x' & y' \end{vmatrix} = 0, \text{ respectively.}$$

5. Show that the solution of equations (1) and (2) is given by

$$\begin{vmatrix} X & \\ c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = \begin{vmatrix} Y & \\ a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = \begin{vmatrix} -1 & \\ a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

A determinant is always square and the vertical lines imply an operation just as +, -, x and \div do. We now define another term, a matrix. A matrix is a rectangular array of numbers. No operation upon these numbers is implied, as in the case of a determinant, but the usual arithmetic operations are defined for matrices. Some device is usually used to contain the numbers of a matrix. For example, we write

$$(X \ Y) \quad \text{or} \quad \| X \ Y \|,$$

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \quad \text{or} \quad \left\| \begin{array}{cc} a_1 & b_1 \\ a_2 & b_2 \end{array} \right\|, \text{ and so on.}$$

The first matrix is simply a pair of numbers used to locate a point on a plane; the second matrix is that of the coefficients of X and Y in equations (1) and (2).

Addition of matrices is defined as the addition of elements in corresponding positions. This implies that the matrices must have the same number of rows and columns. Thus,

$$\begin{aligned} \| 2 \quad 7 \| + \| 3 \quad 5 \| &= \| 5 \quad 12 \| \\ \text{and} \quad \| 3 \quad -1 \| + \| -8 \quad -6 \| &= \| -5 \quad -7 \| . \end{aligned}$$

Subtraction is similarly defined.

Multiplication is defined in a row-by-column manner. The element in the i,j-th position of the resulting matrix is the sum of products of similarly-located elements in the i-th row of the first matrix and the j-th column of the second. For example,

$$\begin{aligned} \left\| \begin{array}{cc} a & b \\ c & d \end{array} \right\| \left\| \begin{array}{cc} e & f \\ g & h \end{array} \right\| &= \left\| \begin{array}{cc} ae + bg & af + bh \\ ce + dg & cf + dh \end{array} \right\| \\ \text{and} \quad \left\| \begin{array}{cc} a_1 & b_1 \\ a_2 & b_2 \end{array} \right\| \left\| \begin{array}{c} X \\ Y \end{array} \right\| &= \left\| \begin{array}{c} a_1X + b_1Y \\ a_2X + b_2Y \end{array} \right\| \end{aligned}$$

In general

$$\left\| a_{ij} \right\| \left\| b_{ij} \right\| = \left\| \sum_j a_{ij} b_{jk} \right\| = \left\| c_{ik} \right\| .$$

When two matrices are multiplied, the number of columns in the first must equal the number of rows in the second. Neither matrix need be square. The resulting matrix will have as many rows as the first and columns as the second.

Division is defined for square matrices as the inverse of multiplication so that

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{-1} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

where the superscript -1 is read as "inverse". We can solve for $\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{-1}$ as follows. Set

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{-1} = \begin{vmatrix} x & y \\ z & w \end{vmatrix}. \quad \text{Then} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} x & y \\ z & w \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

by definition, so that $ax + bz = 1$, $ay + bw = 0$, $cx + dz = 0$, and $cy + dw = 1$. Since these equations are really two pairs, the solution is easy. We have

$$x = \frac{-\begin{vmatrix} -1 & b \\ 0 & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \quad z = \frac{-\begin{vmatrix} a & -1 \\ c & 0 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \quad y = \frac{-\begin{vmatrix} 0 & b \\ -1 & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \quad w = \frac{-\begin{vmatrix} a & 0 \\ c & -1 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}.$$

$$= \frac{d}{ad - bc}, \quad = \frac{-c}{ad - bc}, \quad = \frac{-b}{ad - bc}, \quad = \frac{a}{ad - bc}.$$

The inverse of $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ is $\begin{vmatrix} \frac{d}{ad - bc} & \frac{-b}{ad - bc} \\ \frac{-c}{ad - bc} & \frac{a}{ad - bc} \end{vmatrix}.$

Notice that each denominator is the determinant of the original matrix and that each numerator involves a single element of it. The letters are now in the reverse order; where they went clockwise a, b, d, c , they now go counter-clockwise a, b, d, c with some sign changes.

Systems of simultaneous equations are sometimes written using matrix notation. E.g.

$$\begin{vmatrix} n & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \dots & \sum X_{1i} X_{ki} \\ \sum X_{2i} & \sum X_{2i} X_{1i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki} X_{1i} & \sum X_{ki} X_{2i} & \dots & \sum X_{ki}^2 \end{vmatrix} \times \begin{vmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{vmatrix} = \begin{vmatrix} \sum Y_i \\ \sum X_{1i} Y_i \\ \sum X_{2i} Y_i \\ \vdots \\ \sum X_{ki} Y_i \end{vmatrix}$$

When the two matrices on the left are multiplied, they give a $(k+1) \times 1$ matrix as is the one on the right. Corresponding elements are now equated giving rise to the system required for the general k -variate analysis. Note the symmetry of the matrix of coefficients; if the top row and left column were crossed out, the remaining matrix would be the covariance matrix.

Problems:

1. Add i. $\begin{vmatrix} 2 & 7 \end{vmatrix}$ and $\begin{vmatrix} 9 & 9 \end{vmatrix}$

ii. $\begin{vmatrix} 3 & 0 \\ 4 & 8 \end{vmatrix}$ and $\begin{vmatrix} -8 & 6 \\ 3 & -2 \end{vmatrix}$

2. Subtract i. $\begin{vmatrix} 3 & 8 \end{vmatrix}$ from $\begin{vmatrix} 11 & -6 \end{vmatrix}$

ii. $\begin{vmatrix} 3 & 0 \\ 4 & 8 \end{vmatrix}$ from $\begin{vmatrix} -8 & 6 \\ 3 & -2 \end{vmatrix}$

3. i. From your experience with questions 1 and 2, how do you think

a $\begin{vmatrix} x & y \\ z & w \end{vmatrix}$ should be defined where a is a constant multiplier?

ii. Using your definition above, complete the multiplication and find the determinant of the result.

Is the value of this determinant equal to $a \begin{vmatrix} x & y \\ z & w \end{vmatrix}$?

4. Show that the inverse of $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ is both a right and a left inverse.

5. Find the inverse of $\begin{vmatrix} 3 & 5 \\ 5 & 9 \end{vmatrix}$, $\begin{vmatrix} 4 & -3 \\ -3 & 7 \end{vmatrix}$. Check that your inverses are both right and left inverses.

6. Multiply i. $\begin{vmatrix} 3 & 0 \\ 1 & 5 \end{vmatrix}$ by $\begin{vmatrix} 4 & 5 \\ 5 & 8 \end{vmatrix}$

ii. $\begin{vmatrix} 4 & 7 & 2 \\ 3 & 9 & 1 \end{vmatrix}$ by $\begin{vmatrix} 6 \\ -7 \\ 2 \end{vmatrix}$

How many ways can the multiplication be done in i.? In ii.? Do they lead to identical or symmetrical results?

7. Write the equations of problem 3, page 3, in matrix notation. Why is the matrix of coefficients not symmetric? If this matrix is denoted by A, then

$$A \begin{vmatrix} \hat{\alpha} \\ \hat{\beta} \end{vmatrix} = \begin{vmatrix} \bar{y} \\ \Sigma X_i Y_i \end{vmatrix}. \quad \text{Show that} \quad \begin{vmatrix} \hat{\alpha} \\ \hat{\beta} \end{vmatrix} = A^{-1} \begin{vmatrix} \bar{y} \\ \Sigma X_i Y_i \end{vmatrix}.$$

Consider a completely random design in which there are only two treatments and n observations are made on each. The model is

$$X_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, \quad j = 1, \dots, n$$

Following our rules for writing least squares equations from which to estimate μ , τ_1 and τ_2 , we have

$$\sum_{i,j} (X_{ij} - \hat{\mu} - \hat{\tau}_i) = 0 \quad (8)$$

$$\sum_j (X_{1j} - \hat{\mu} - \hat{\tau}_1) = 0 \quad (9)$$

$$\sum_j (X_{2j} - \hat{\mu} - \hat{\tau}_2) = 0 \quad (10)$$

Notice that rule iii. which calls for a summation sign does not tell us what subscripts we sum over. This is easily answered. In equation (8), the $\hat{\mu}$ equation, we sum over i and j because every residual $(X_{ij} - \hat{\mu} - \hat{\tau}_i)$ is concerned with $\hat{\mu}$. In equation (9), the $\hat{\tau}_1$ equation, we sum over j only since only residuals containing $\hat{\tau}_1$ supply any information about $\hat{\tau}_1$. Consequently, we don't make use of residuals $(X_{2j} - \hat{\mu} - \hat{\tau}_2)$ so don't sum over i . Similarly for the $\hat{\tau}_2$ equation. You may also argue from rule ii. that the coefficient of $\hat{\tau}_1$ is zero for residuals $(X_{2j} - \hat{\mu} - \hat{\tau}_2)$ so such terms disappear from the equation. In other words, we don't sum over i .

Equations (8), (9) and (10) reduce to

$$\sum_{i,j} X_{ij} - 2n\hat{\mu} - n\hat{\tau}_1 - n\hat{\tau}_2 = 0 \quad (11)$$

$$\sum_j X_{1j} - n\hat{\mu} - n\hat{\tau}_1 = 0 \quad (12)$$

$$\sum_j X_{2j} - n\hat{\mu} - n\hat{\tau}_2 = 0 \quad (13)$$

From equations (12) and (13),

$$(\hat{\mu} + \hat{\tau}_1) = \bar{x}_1.$$

and

$$(\hat{\mu} + \hat{\tau}_2) = \bar{x}_2.$$

It is immediately obvious that equation (11) is the sum of equations (12) and (13) and is useless in further solving our equations. We need one more independent equation in order to obtain a solution. We have already seen that this is

$$(\hat{\tau}_1 + \hat{\tau}_2) = 0 \quad (14)$$

Equation (11) now becomes

$$\sum_{i,j} X_{ij} - 2n\hat{\mu} = 0$$

from which

$$\hat{\mu} = \bar{x}.$$

Now $\hat{\tau}_1 = \bar{x}_1 - \bar{x}$

and

$$\hat{\tau}_2 = \bar{x}_2 - \bar{x}.$$

Problems:

1. Write out the least squares equations for
 - i. a completely random design with t treatments, each replicated n_i times, $i = 1, \dots, t$.
 - ii. a randomized complete block design with r replicates and t treatments.
 - iii. a latin square design of side r .
2. Show that the $\hat{\mu}$ equation is the sum of some or all of the other equations in all cases.
3. How many independent equations do you have for each case?
4. What additional equation or equations are needed in each of the three cases?

Equations such as 11, 12, 13, 14 can be written in matrix notation and solved by determinantal methods. Thus, the equations may be written

$$\begin{vmatrix} 2n & n & n \\ n & n & 0 \\ n & 0 & n \\ 0 & 1 & 1 \end{vmatrix} \begin{vmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{vmatrix} = \begin{vmatrix} \sum_{i,j} X_{ij} \\ \sum_j X_{1j} \\ \sum_j X_{2j} \\ 0 \end{vmatrix} \quad (15)$$

Since determinants have not been defined for other than square arrays, we cannot have a determinant for the 4x3 matrix. We have already seen that the first three equations were not independent, that one was the sum of the other two, and was useless in terms of solving for the three unknowns. Any one of these equations may be safely omitted. Thus, we have

$$\begin{vmatrix} n & n & 0 \\ n & 0 & n \\ 0 & 1 & 1 \end{vmatrix} \begin{vmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{vmatrix} = \begin{vmatrix} \sum_j X_{1j} \\ \sum_j X_{2j} \\ 0 \end{vmatrix} \quad (16)$$

as a sufficient set of equations.

We have already seen that the solution of these equations is $\hat{\mu} = \bar{x}$, $\hat{\tau}_1 = \bar{x}_1 - \bar{x}$, $\hat{\tau}_2 = \bar{x}_2 - \bar{x}$. In problem 7, page 6, we observed, for a particular set of equations, that a solution of

$$A \begin{vmatrix} \hat{\alpha} \\ \hat{\beta} \end{vmatrix} = \begin{vmatrix} \bar{y} \\ \sum_i X_i Y_i \end{vmatrix} \text{ is given by } \begin{vmatrix} \hat{\alpha} \\ \hat{\beta} \end{vmatrix} = A^{-1} \begin{vmatrix} \bar{y} \\ \sum_i X_i Y_i \end{vmatrix}$$

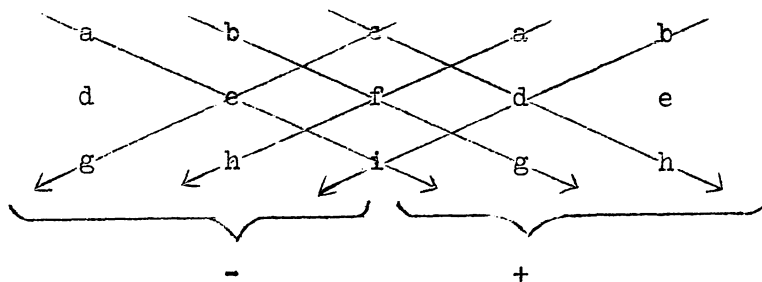
where A is the matrix of coefficients of the unknowns, $\hat{\alpha}$ and $\hat{\beta}$.

This method of solution is general provided A is square and its determinant is not zero. This turns out to mean that the equations must be independent, none being a linear combination of any of the others. For a 3x3 matrix, the determinant is defined as follows:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \quad (17)$$

$$= aei - afh - bdi + bfg + cdh - ceg.$$

This may also be computed as indicated in the following sketch:



or $aei + bfg + cdh - ceg - afh - bdi$, as before. The expansion method, but not the sketched method, is a general one. We have

$$\begin{vmatrix} n & n & 0 \\ n & 0 & n \\ 0 & 1 & 1 \end{vmatrix} = 0 - n^2 - n^2 + 0 + 0 - 0 = -2n^2$$

as the determinant of the 3×3 matrix in (16).

Some useful general rules for handling determinants follow.

1. The value of a determinant is not altered by changing rows into columns and columns into rows.
2. If two rows or two columns are interchanged, the determinant is changed in sign only.
3. If two rows or two columns are identical, the determinant is zero.
4. If each element in any row, or any column, is multiplied by the same factor, then the determinant is multiplied by that factor.
5. Any row or column can be increased or decreased by adding multiples of any of the other rows or columns. The value of the determinant is unaltered.

Problems:

1. Find the values of

$$\text{i)} \begin{vmatrix} 1 & 1 & 1 \\ 3 & 0 & 3 \\ 2 & 8 & 9 \end{vmatrix}, \quad \text{ii)} \begin{vmatrix} 1 & 2 & 3 \\ 0 & 4 & 6 \\ 0 & 1 & 2 \end{vmatrix}, \quad \text{iii)} \begin{vmatrix} .7 & 3 & 1 \\ .6 & 6 & 0 \\ .3 & 3 & 1 \end{vmatrix}, \quad \text{iv)} \begin{vmatrix} 1 & 3 & 2 \\ 4 & 0 & 4 \\ 2 & 6 & 4 \end{vmatrix}$$

If we write a set of three linear equations as

$$\begin{aligned} a_1x + b_1y + c_1z + d_1 &= 0 \\ a_2x + b_2y + c_2z + d_2 &= 0 \\ a_3x + b_3y + c_3z + d_3 &= 0, \end{aligned} \tag{18}$$

then their solution is given by

$$\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \frac{-1}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \tag{19}$$

Problems:

1. Show that (19) may be written symmetrically as

$$\begin{array}{c} \text{x} \\ \hline \begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix} \end{array} = \begin{array}{c} \text{-y} \\ \hline \begin{vmatrix} a_1 & c_1 & d_1 \\ a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \end{vmatrix} \end{array} = \begin{array}{c} \text{z} \\ \hline \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} \end{array} = \begin{array}{c} \text{-1} \\ \hline \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \end{array}$$

2. Show that

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} \begin{vmatrix} e & f \\ h & i \end{vmatrix} & -\begin{vmatrix} b & c \\ h & i \end{vmatrix} & \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ -\begin{vmatrix} d & f \\ g & i \end{vmatrix} & \begin{vmatrix} a & c \\ g & i \end{vmatrix} & -\begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ \begin{vmatrix} d & e \\ g & h \end{vmatrix} & -\begin{vmatrix} a & b \\ g & h \end{vmatrix} & \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{vmatrix} = \begin{vmatrix} |D| & 0 & 0 \\ 0 & |D| & 0 \\ 0 & 0 & |D| \end{vmatrix}$$

where $|D|$ is the determinant of $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$.

3. How would you describe the formation of the matrix of determinants in terms of the original matrix?

The Relation Between Experimental Design and Multiple Regression

The general multiple regression model is usually written as

$$Y_i = \mu + \sum_{\alpha} \beta_{\alpha} X_{\alpha i} + \epsilon_i$$

or as

$$y_i = \sum_{\alpha} \beta_{\alpha} x_{\alpha i} + \epsilon_i. \quad (15)$$

Here the $x_{\alpha i}$'s are observed parameters in the population mean whereas the β_{α} 's are unknown and present a problem of estimation.

In an experimental design model, the $x_{\alpha i}$'s take on only the values 0 or 1 so are not generally shown, the β_{α} 's are written as ρ 's, τ 's, etc. and certain relations usually exist among them. For example consider a randomized block experiment with two treatments and two replicates. The usual model is

$$X_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}, \quad i = 1, 2, \quad j = 1, 2. \quad (16)$$

However, it may be written as

$$Y_{ij} = \mu + X_{1ij}\rho_1 + X_{2ij}\rho_2 + X_{3ij}\tau_1 + X_{4ij}\tau_2 + \epsilon_{ij}$$

where $X_{1ij} = 1$ for $i = 1$ and zero otherwise, $X_{2ij} = 1$ for $i = 2$ and zero otherwise, $X_{3ij} = 1$ for $j = 1$ and zero otherwise, and $X_{4ij} = 1$ for $j = 2$ and zero otherwise. Thus, we have

$$\begin{aligned} Y_{11} &= \mu + \rho_1 + \tau_1 + \epsilon_{11} \\ Y_{12} &= \mu + \rho_1 + \tau_2 + \epsilon_{12} \\ Y_{21} &= \mu + \rho_2 + \tau_1 + \epsilon_{11} \\ Y_{22} &= \mu + \rho_2 + \tau_2 + \epsilon_{11}, \end{aligned} \quad (17)$$

a set of equations more easily written as (16).

Problems:

1. Substitute $\rho_1 = -\rho_2$, $\tau_1 = -\tau_2$ and show that the reduction attributable to regression is $\hat{\rho}_1 \sum (X_{1ij} - X_{2ij})y_{ij} + \hat{\tau}_1 \sum (X_{3ij} - X_{4ij})y_{ij}$ and that this is equivalent to the anova reduction.
2. Consider the multiple regression equation (15). Write the normal equations and their solution in matrix notation. Relate this to Snedecor's Gauss multiplier procedure in section 13.12.

Let $\|a_{ij}\|$ denote a square matrix with determinant $\neq 0$. If we strike out the i -th row and j -th column, and take the determinant of this new matrix, we have the minor of a_{ij} , say M_{ij} . Thus, in problem 2, page 1, $\begin{vmatrix} b & c \\ h & i \end{vmatrix}$ is the minor of d . The cofactor of a_{ij} , denoted by A_{ij} , is defined as $(-1)^{i+j}M_{ij}$. Thus, the cofactor of d is $(-1)^{1+2} \begin{vmatrix} b & c \\ h & i \end{vmatrix} = - \begin{vmatrix} b & c \\ h & i \end{vmatrix}$. The matrix of cofactors is written as $\|A_{ij}\|$. It can be shown that $\|a_{ij}\| \cdot \|A_{ji}\| = \begin{vmatrix} D & \cdots & 0 \\ \vdots & D & \vdots \\ 0 & \cdots & D \end{vmatrix}$ where D is the determinant of the matrix $\|a_{ij}\|$. Note the change of subscripts in $\|A_{ji}\|$. This means we have interchanged rows and columns. The matrix $\|A_{ji}\|$ is also called the transpose of $\|A_{ij}\|$.

We now have a method for computing the inverse of a matrix: form the matrix of cofactors; divide each element by the determinant of the matrix; the transpose of this matrix is the inverse of $\|a_{ij}\|$. The inverse of a matrix with zero determinant is not defined.

Problem: Find the inverses of

$$\begin{array}{lll} \text{i)} & \left\| \begin{array}{ccc} 1 & 1 & 1 \\ 3 & 0 & 3 \\ 2 & 8 & 9 \end{array} \right\|, & \text{ii)} & \left\| \begin{array}{ccc} 1 & 2 & 3 \\ 0 & 4 & 6 \\ 0 & 1 & 2 \end{array} \right\|, & \text{iii)} & \left\| \begin{array}{ccc} 1 & 3 & 2 \\ 4 & 0 & 4 \\ 2 & 6 & 4 \end{array} \right\| \end{array}$$

by use of the cofactor matrix.